УДК 519.95
А.Н.Зинченко
Новокаховский гуманитарный институт

ОПТИМАЛЬНАЯ ФИЛЬТРАЦИЯ В ЗАДАЧАХ КОНТРОЛЯ КАЧЕСТВА ТЕКСТИЛЬНЫХ МАТЕРИАЛОВ

В работе решены задачи автоматизированного контроля качества текстильных материалов, решение которых позволяет в автоматическом режиме определять дефекты ткани по ее изображению.

Ключевые слова: *оптимальная фильтрация, автоматизированный контроль.*

Постановка проблемы. Задачи автоматизированного контроля качества вырабатываемой ткани, решение которых позволит в автоматическом режиме определять дефекты ткани по ее изображению, в настоящее время являются достаточно актуальными для предприятий текстильной промышленности, поскольку позволяют существенно улучшить качество продукции и значительно увеличить ее конкурентоспособность, что является решающим фактором в условиях членства Украины в ВТО.

Состояние изученности проблемы. Проблемами оптимизации фильтрации в задачах контроля качества занимались такие зарубежные и отечественные ученые, как Р. Богнер, Э. Оппенгейм, Г. Шлихт, С.М. Арбузов, Д. Ватолин, А.Н. Писаревский, Е.Б. Ратушняк, А.Б. Сергиенко, М. Смирнов, Соловьева, А.И. Солонина, Д.А. Улахович, В. Юкин. Несмотря на это методологические аспекты фильтрации при проведении контроля над качеством именно-текстильных материалов на сегодняшний день изучены недостаточно.

Задание и методика исследований. Задание проведенной работы заключалось в исследовании схемы „технического зрения” на конвейерной линии, проверке точности существующих методов фильтрации в задачах контроля качества и разработке новых методов, используемых только текстильными предприятиями. Исследование проводилось при помощи абстрактно – логического, монографического, графического и аналитического методов.

Результаты исследований. Обобщенная функциональная схема системы технического зрения (СТЗ) приведена на рис.1. Изображение объекта через оптический блок передается на преобразователь „свет-сигнал” (обычно матричный), электрический сигнал которого в устройстве первичной обработки изображения усиливается и запоминается. Устройство анализа изображения (вторичной обработки) служит для выделения и опознавания объекта, определения его координат и положения. При необходимости обработанная информация об объекте высвечивается на устройстве визуального контроля. На основе полученной информации контроллер связи с исполнительным механизмом вырабатывает управляющие сигналы, приводящие в действие исполнительные механизмы, осуществляющие целенаправленное воздействие на объект.

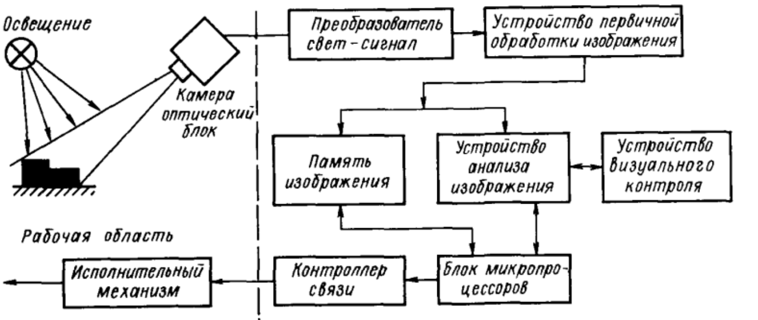

Рис. 1.Общая функциональная схема СТЗ

Чувствительным элементом в рассмотренной схеме является оптический блок с преобразователем „свет-сигнал”. Обычно он реализуется с помощью телекамеры. Существуют варианты СТЗ с различной разрешающей способностью (64 - 256) и уровнями квантования видеосигнала (1 - 256). В настоящее время выделяют несколько поколений СТЗ. Системы первого поколения оснащены микро-ЭВМ и телекамерой и работают по бинарному принципу, т. е.

представляют все точки визируемой зоны как светлые и темные. Такое представление налагает жесткие требования на условия освещения объекта, поскольку изображение должно быть контрастным. Системы второго поколения способны анализировать уровни освещенности, т. е. правильно определять границы деталей и некоторые существенные особенности.

Назначение устройства первичной обработки состоит в уменьшении общего времени обработки изображения. При этом в большинстве случаев осуществляется фиксация и представление изображения в виде характерных точек (определение центра объекта, нахождение его важных характерных особенностей: углов, расстояний от центра до края и т. д.). Таким образом, на первом этапе осуществляется непосредственная обработка и представление информации в виде, удобном для дальнейших стандартных преобразований, которые обычно выполняют на универсальной ЭВМ. Сама же предварительная обработка учитывает специфику объекта, поэтому устройство первичной обработки, как правило, является специализированным, реализующим требуемые алгоритмы. Наличие в современных СТЗ устройства первичной обработки обусловлено требованиями высокой скорости обработки данных (не более 150 мс на изображение), технологической гибкости и высокой разрешающей способности. Отсутствие этого устройства в системе увеличивает общее время обработки.

Устройство вторичной обработки осуществляет распознавание изображения, т. е. по его характерным признакам определяет, к какому классу принадлежит визируемый объект. На вход устройства поступает код изображения, полученный на этапе первичной обработки. На выходе образуются управляющие команды, преобразуемые в дальнейшем контроллером связи в сигналы, определяющие перемещения и действия манипулятора. Процесс удешевления аппаратуры позволил для вторичной обработки использовать микропроцессоры, а также мини- и микро-ЭВМ. Для обучения роботов и наблюдения за рабочей ситуацией иногда в СТЗ вводят устройство визуального контроля.

Адаптивные устройства обработки данных отличаются наличием определенной связи параметров передаточной функции с параметрами входных, выходных, ожидаемых, прогнозируемых и прочих дополнительных сигналов или с параметрами их статистических соотношений. В простейшем случае, адаптивное устройство содержит программируемый фильтр обработки данных и блок (алгоритм) адаптации, который на основании определенной программы анализа входных, выходных и прочих дополнительных данных вырабатывает сигнал управления параметрами программируемого фильтра. Импульсная характеристика адаптивных систем также может иметь как конечный, так и бесконечный характер.

Как правило, адаптивные устройства выполняются узко целевого функционального назначения под определенные типы сигналов. Внутренняя структура адаптивных систем и алгоритм адаптации практически полностью регламентируются функциональным назначением и определенным минимальным объемом исходной априорной информации о характере входных данных и их статистических и информационных параметрах. Это порождает многообразие подходов при разработке систем, существенно затрудняет их классификацию и разработку общих теоретических положений.

Фильтрация шумов и сглаживание изображений.

В настоящее время для фильтрации шумов все чаще используют локальные нелинейные фильтры. Простейшим локальным нелинейным фильтром является медианный фильтр [1], выход которого определяется как медиана элементов, попавших в апертуру фильтра. Более эффективными являются комбинированные фильтры [2] в которых к обрабатываемому изображению последовательно применяются несколько алгоритмов фильтрации.

Степень сглаживания изображения напрямую зависит от размера апертуры локального фильтра: при малом размере апертуры лучше сохраняются „контрастные” детали изображения, но шум сглаживается в малой степени. При большом размере апертуры наблюдается обратная картина. Это противоречие, в определенной степени, удается разрешить в локальных фильтрах с адаптацией размера апертуры [5]. В этих фильтрах большие размеры апертуры используются в монотонных областях обрабатываемого изображения (обеспечивая тем самым лучшее подавление шума), а маленькие размеры – вблизи границ, ребер изображения (обеспечивая сохранение этих контрастных деталей).

Постановка задачи адаптивной обработки сигналов.

Одним из эффективных путей решения класса задач обработки сигналов в условиях априорной неопределенности может быть применение разнообразных методов адаптации. В этом случае задача решается так же, как при отсутствии неопределенности, а затем в синтезированные алгоритмы обработки сигналов вместо неизвестных параметров вставляются их оценки (в статистическом смысле), полученные по входным выборкам. Естественно предполагать, что эффективность описанного алгоритма будет ниже, чем при наличии полной априорной информации, т. к. оценки неизвестных параметров вычисляются с определенной ошибкой. Для оценки степени достижения требуемого качества адаптации обычно вводится функционал качества (функция качества, рабочая функция, стоимостная функция), зависящий как от входного сигнала, так и от параметров системы ЦОС. Достижение экстремума данного функционала (локального или глобального) является целью функционирования адаптивной системы [6].

В большинстве случаев, особенно при обработке объемных выборок, оказывается, что определение оценок параметров системы и их функциональное преобразование связаны со значительными вычислительными трудностями. Поэтому используется другой метод адаптации, заключающийся в том, что оптимальное решение, соответствующее экстремуму функционала качества, Достигается путем последовательных приближений. Сначала задача решается при полной определенности. Шаги последовательного приближения определяются по вычисляемым значениям детерминированного функционала качества и его производных. Таким образом, получается детерминированный итерационный алгоритм. Поскольку в условиях априорной неопределенности Функционал качества зависит от нескольких неизвестных параметров и не может быть вычислен непосредственно, в синтезированных алгоритмах его значения (и значения его производных) заменяют их оценками. Получающийся, при этом, алгоритм называется итерационным стохастическим и соответствует теореме о разделении.

Синтез детерминированных итерационных алгоритмов базируется на результатах математической теории оптимизации как без ограничений, так и с ограничениями. В процессе адаптации основное внимание уделяется решению задач оптимизации с квадратичной функцией качества. Это связано с тем, что ряд основных параметров систем адаптивной обработки сигналов выражается в виде квадратичных форм, сочетающихся с простотой и наглядностью получаемых при этом результатов. Кроме того, из математики известно [7], что не квадратичные функции качества можно аппроксимировать квадратичной зависимостью, раскладывая их в ряд Тейлора вблизи экстремальной точки и ограничиваясь тремя членами разложения.

В общем случае графически функционал качества представляет собой гиперповерхность в многомерном пространстве, связанным с количеством независимых варьируемых параметров адаптивной системы. Данная гиперповерхность обладает следующим важным свойством: если случайные сигналы являются стационарными и их вероятностные характеристики инвариантны относительно временных сдвигов, то эта гиперповерхность фиксирована и остается неподвижной в своей системе координат. В этом случае процесс адаптации заключается в движении, начиная с некоторой начальной точки, в направлении экстремума и в удержании среднеквадратичного значения сигнала ошибки вблизи этой точки.

Если сигналы нестационарны и их статистические свойства меняются во времени, то можно считать описанную гиперповерхность „размытой”, изменяющей свою форму и местоположение относительно введенной системы координат. В этом случае процесс адаптации состоит не только в движении к точке экстремума, но и в слежении за этой точкой, поскольку она меняет свое положение в пространстве.

Оптимальное нерекурсивное оценивание.

Принципы оптимального линейного оценивания являются фундаментальными при любом рассмотрении адаптивных систем обработки сигналов и, в частности, адаптивных фильтров. Процесс адаптивной фильтрации включает два этапа проведения оценивания:

- оценивание искомого выходного сигнала фильтра;
- оценивание коэффициентов фильтра (отсчетов импульсной характеристики), необходимых для достижения поставленной цели.

Второй этап необходим из-за априорной неопределенности входного сигнала, подвергшегося воздействию шумовой помехи.

Простейшей и наиболее широко распространенной адаптивной структурой является КИХ - фильтр с регулируемыми коэффициентами. Схема этого фильтра представлена на рис. 2. Иногда его называют адаптивным линейным сумматором [4].

На выходе фильтра необходимо получить оценку y(n), максимально соответствующую (в смысле выбранного критерия качества) незашумленному сигналу y(n).

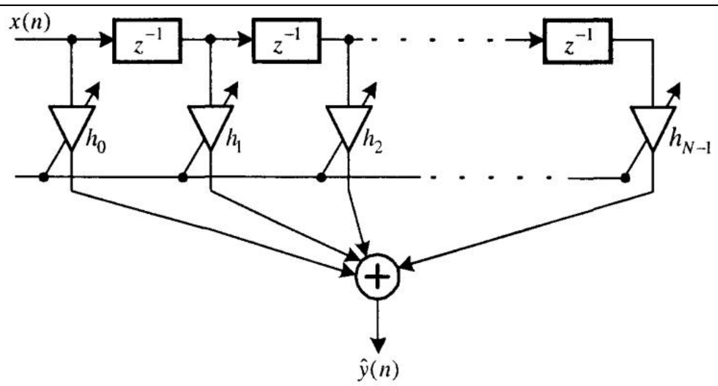

Рис. 2. Адаптивный КИХ – фильтр

Синтез устройства оценки на базе адаптивного КИХ-фильтра существенно зависит от определения стоимостной функции, в соответствии с которым качество оценивания характеризуется разностью между выходным сигналом устройства оценки и истинным значением, подлежащим оцениванию:

$$e(n) = y(n) - \hat{y}(n), \tag{1}$$

где e(n) - ошибка оценивания;y(n) - оцениваемый случайный сигнал; y(n) - его статистическая оценка.

В нашем случае оценка y(n) является линейной функцией последовательности входных отсчетов x(n) и коэффициентов фильтра h_n (n = 0,1,..., N -1). Последовательность отсчетов x(n) в общем виде можно представить как сигнал y(n), искаженный аддитивным белым шумом ν(n) с дисперсией σ:

$$x(n) = y(n) + v(n)\,. \tag{2}$$

Наиболее часто используемым при проведении оптимального оценивания y(n) является уже упоминавшийся метод наименьших квадратов (МНК). При этом среднеквадратичная ошибка определяется как

$$E\left[e^2(n)\right] = E\left\{\left[y(n) - \hat{y}(n)\right]^2\right\}, \tag{3}$$

где E [e^2(n)] - оператор математического ожидания.

Среднеквадратичная ошибка минимизируется относительно весовых коэффициентов КИХ - фильтра для получения оптимального оценивания по критерию МНК.

В нерекурсивном фильтре в соответствии с его разностным уравнением выходная оценка y(n) является конечным линейным полиномом:

$$\hat{y}(n) = \sum_{k=0}^{N-1} h_k x(n-k)\,. \tag{4}$$

Выражение (4) можно переписать в векторно-матричной системе обозначений

$$\hat{y}(n) = X^T(n) \cdot H = H^T X(n)\,, \tag{5}$$

где:

$$X(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-N+1) \end{bmatrix}$$

— вектор-столбец входного сигнала;

$$H = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{N-1} \end{bmatrix}$$

— вектор-столбец коэффициентов фильтра.

Тогда функция среднеквадратичной ошибки (3) принимает вид:

$$E\left[e^2(n)\right]=E\left[y(n)-H^T X(n)\right]^2 \tag{6}$$

Это выражение описывает стандартную поверхность гиперпараболоид в (N +1) -мерном пространстве с единственным минимумом. Дифференцирование (6) по H дает

$$\frac{\partial E\left[e^2(n)\right]}{\partial H^T}=-2E\left\{\left[y(n)-H^T X(n)\right]X^T(n)\right\} \tag{7}$$

Допуская, что (7) равно нулю, получаем

$$E\left\{\left[y(n)-H^T X(n)\right]X^T(n)\right\}=0 \tag{8}$$

или, что то же самое,

$$E\left[y(n)X^T(n)\right]=E\left[H^T X(n)X^T(n)\right] \tag{9}$$

Полагая, что вектор коэффициентов H и вектор входного сигнала X(n) некоррелированы, получаем

$$E\left[y(n)X^T(n)\right]=H_0^T\cdot E\left[X(n)X^T(n)\right], \tag{10}$$

где H_o - вектор оптимальных коэффициентов КИХ - фильтра, обеспечивающий минимум выражению (6).

Члены математических ожиданий в выражении (10) можно определить следующим образом:

$$R=E\left[X(n)X^T(n)\right]$$

P - E[y(n) * X(n)] - вектор взаимной корреляции между оцениваемым сигналом и отсчетами входной последовательности размером Nx. С у четом введенных обозначений (10) можно переписать в виде:

$$P^T=H_0^T R. \tag{11}$$

Уравнение (11) является известным матричным уравнением Винера— Хопфа, которое дает оптимальное (по критерию МНК) решение для коэффициентов КИХ - фильтра

$$H_0=R^{-1}P. \tag{12}$$

Выражение (12) получено из (11) с учетом симметричности корреляционной матрицы R, для которой

$$\left[R^{-1}\right]^T=R^{-1}.$$

Винеровская оценка (12) по существу является одношаговым блочным процессом, который подходит для конечной выборки (блока) данных. В случае нестационарности входного сигнала обновление матриц R и P должно происходить на каждом временном шаге.

Получим остаточную среднеквадратичную ошибку оценивания [4], используя оптимальный вектор коэффициентов H_o. Из соотношения (8) можно получить выражение

$$E\left[e(n)X(n)\right]=0. \tag{13}$$

Преобразуем формулу для среднеквадратичной ошибки с учетом (13) и вычисленного вектора H_o:

$$\begin{aligned}E\left[e^2(n)\right]&=E\left\{e(n)\left[y(n)-H_0{}^T X(n)\right]\right\}=\\&=E\left[e(n)\cdot y(n)\right]-E\left\{\left[y(n)-H_0{}^T X(n)\right]\cdot y(n)\right\}=\\&=E\left[y^2(n)\right]-H_0{}^T E\left[y(n)\ \ X(n)\right]=E\left[y^2(n)\right]-H_0{}^T P.\end{aligned} \tag{14}$$

Формула (14) позволяет вычислять остаточную среднеквадратичную ошибку при известном полезном сигнале y(n) и найденных векторах H_o и P.

Рекуррентные алгоритмы адаптации.

Рассмотренная ранее винеровская оценка коэффициентов КИХ - фильтра требует полного пересчета всех членов авто- и взаимокорреляционных матриц для каждой новой выборки, что с вычислительной точки зрения нерационально. Если иметь дело с продолжительным (теоретически бесконечным) рядом отсчетов входного сигнала, значительно более удобными являются рекуррентные алгоритмы получения оценок, вносящие коррекцию на каждом шаге итерационного процесса.

Калмановское оценивание случайного сигнала.

По существу, калмановское оценивание реализует рекурсивную процедуру адаптации, основанную на авторегрессионной модели процесса генерирования сигнала. Если входной сигнал x(n) является случайным и марковским, то его можно представить в виде выхода ЛДС первого порядка, возбуждаемой белым шумом w(n) с нулевым средним и дисперсией a_w. Модель генерирования сигнала описывается разностным уравнением 1-го порядка [6]:

$$x(n) = ax(n-1) + w(n-1) \tag{15}$$

Структурная схема устройства, соответствующая уравнению (1), представлена на рис. 3.

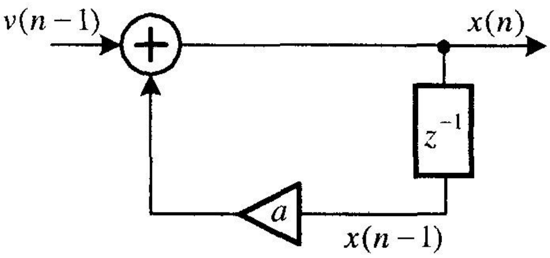

Рис. 3. Устройство генерирования случайного сигнала

Предполагается, что после прохождения канала связи сигнал x(n) претерпел амплитудные изменения, описываемые постоянным коэффициентом c, и на него воздействовал аддитивный белый шум v(n) с нулевым средним и дисперсией a. Модель воздействия канала на сигнал описывается простым уравнением

$$y(n) = cx(n) + v(n)\,. \tag{16}$$

Соответствующая ей структурная схема представлена на рис. 4.

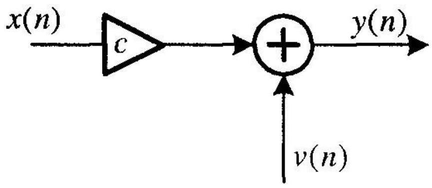

Рис. 4. Модель прохождения сигнала по каналу связи

Зашумленный сигнал y(n) поступает на вход синтезируемого адаптивного калмановского фильтра. На его выходе необходимо получить рекуррентную оценку x(n), максимально близкую к сигналу x(n) по критерию МНК.

Рекурсивная формула оценки первого порядка имеет вид:

$$\hat{x}(n) = b(n)\hat{x}(n-1) + k(n)\cdot y(n) \tag{17}$$

Следует отметить, что в общем случае коэффициенты b(n) и к(n) зависят от нормированного времени. Обобщенная структурная схема адаптивного оценивателя, реализующего алгоритм (17), представлена на рис. 5.

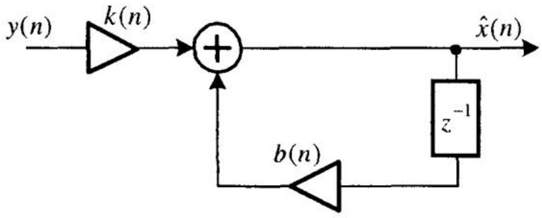

Рис. 5. Обобщенная структура рекурсивного оценивателя первого порядка

$$e(n) = \hat{x}(n) - x(n)\,,$$

$$p(n) = E\left[\hat{x}(n) - x(n)\right]^2. \qquad (18)(19)$$

Введем обозначения:

Выражение (18) называется ошибкой оценки, а (19) - среднеквадратичной ошибкой. Подставляя (17) в (19), получим

$$p(n) = E\left[b(n)\hat{x}(n-1) + k(n)y(n) - x(n)\right]^2$$

Для получения оптимального с точки зрения МНК оценивателя выражение (21) дифференцируется по b(n) и к(n) с последующим приравниванием результатов нулю:

$$\frac{\partial p(n)}{\partial b(n)} = 2E\left\{\left[b(n)\hat{x}(n-1) + k(n)y(n) - x(n)\right]\hat{x}(n-1)\right\} = 0;$$

$$\frac{\partial p(n)}{\partial k(n)} = 2E\left\{\left[b(n)\hat{x}(n-1) + k(n)y(n) - x(n)\right]y(n)\right\} = 0. \qquad (22)(23)$$

Преобразуем уравнение (21):

$$E\left\{\left[b(n)\hat{x}(n-1)\right]\hat{x}(n-1)\right\} = E\left\{-\left[k(n)y(n) - x(n)\right]\hat{x}(n-1)\right\} \qquad (24)$$

После несложных арифметических преобразований из (23) получаем равенство

$$b(n)E\left\{\left[\hat{x}(n-1) - x(n-1) + x(n-1)\right]\hat{x}(n-1)\right\} =$$
$$= E\left\{\left[x(n) - k(n)y(n)\right]\hat{x}(n-1)\right\}. \qquad (25)$$

Подставив в (24) значение y(n) из (16) с учетом обозначения (18), получим:

$$b(n)E\left[e(n-1)\hat{x}(n-1) + x(n-1)\hat{x}(n-1)\right] =$$
$$= E\left\{\left[x(n)\left[1 - ck(n)\right] - k(n)v(n)\right]\hat{x}(n-1)\right\}. \qquad (25)$$

Принцип ортогональности, который минимизирует ошибку, требует некоррелированности ошибки e(n) и оценки x(n-1), а также независимости шума v(n) и x(n-1), что выполняется в рамках сделанных предположений. Это означает выполнение равенств

$$E\left[e(n)\hat{x}(n-1)\right] = 0 \qquad (26)$$

и

$$E\left[v(n)\hat{x}(n-1)\right] = 0 \qquad (27)$$

Тогда уравнение (25) с учетом (26) и (27) примет вид

$$b(n)E\left[x(n-1)\hat{x}(n-1)\right] = \left[1 - ck(n)\right]\cdot E\left[x(n)\hat{x}(n-1)\right] \qquad (28)$$

Подставляя модель генерирования сигнала (15) в (28), получим

$$b(n)E\left[x(n-1)\hat{x}(n-1)\right] =$$
$$= \left[1 - ck(n)\right]E\left[ax(n-1)\hat{x}(n-1) + w(n-1)\hat{x}(n-1)\right] \qquad (29)$$

$$\hat{x}(n-1) = b(n-1)\hat{x}(n-2) + ack(n-1)x(n-2) + \\ +ck(n-1)w(n-2) + k(n-1)v(n-1). \qquad (28.16)$$

Последовательная подстановка (15) в (16), а затем в (17) дает

Умножим обе части равенства (30) на w(n-l) и возьмем математическое ожидание

$$E\left[\hat{x}(n-1)w(n-1)\right] = 0,$$

т. к. шум w(n-1) некоррелирован со всеми членами в правой части (30). Воспользовавшись соотношением (31), преобразуем (29):

$$b(n)E\left[x(n-1)\hat{x}(n-1)\right] = a\left[1-ck(n)\right]E\left[x(n-1)\cdot\hat{x}(n-1)\right], \qquad (31)$$

что приводит к соотношению между коэффициентами b(n) и к(n):

$$b(n) = a\left[1-ck(n)\right]. \qquad (32)$$

Подставив (32) в (17), после несложных преобразований получим

$$\hat{x}(n) = a\hat{x}(n-1) + k(n)\left[y(n) - ac\hat{x}(n-1)\right]. \qquad (33)$$

Уравнение (32) является искомым решением для построения адаптивного рекурсивного оценивателя первого порядка, называемого скалярным фильтром Калмана. Его структурная схема изображена на рис. 6.

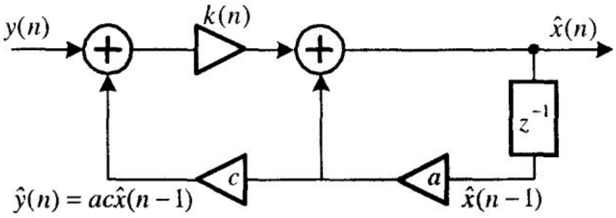

Рис. 6. Скалярный фильтр Калмана

Адаптация в этом устройстве оценки происходит следующим образом. Предыдущая оценка x(n-1) после умножения на коэффициенты а и с предсказывает очередной отсчет зашумленного сигнала y(n). Последний сравнивается с текущим отсчетом y(n). Разница между ними с коэффициентом „доверия” к(n) суммируется с предсказанной оценкой ax(n-1), в результате чего получается текущая оценка x(n).

Нетрудно предположить, что изменяющийся во времени коэффициент „доверия” к(n) должен зависеть от шумовых параметров модели и текущего значения среднеквадратичной ошибки p(n). В [4] получено явное выражение для к(n):

$$k(n) = \frac{c\left[a^2 p(n-1) + \sigma_w^2\right]}{\sigma_v^2 + c^2\sigma_w^2 + c^2a^2p(n-1)} \qquad (34)$$

где

$$p(n) = \frac{1}{c}\sigma_v^2 k(n). \qquad (35)$$

Анализ выражений (28.20) и (28.21) указывает на гибкое (адаптивное) изменение коэффициента к(n) в зависимости от дисперсий действующих шумов σ^2v и σ^2_W, а также величины текущей среднеквадратичной ошибки p(n). Для перехода к векторному фильтру Калмана необходимо перейти к авторегрессионной модели генерирования сигнала более высокого порядка с последующей редукцией к многомерному пространству состояний.

Характеристика итерационных алгоритмов адаптации.

Как было отмечено ранее, синтез итерационных алгоритмов адаптации основывается на результатах математической теории оптимизации как без ограничений, так и с ограничениями. Последние могут быть связаны с частичным заданием структуры схемы ЦОС, с необходимостью фиксации некоторых ее параметров и т. п. В задачах оптимизации с ограничениями выбор промежуточных и конечных решений не может быть произвольным и проводится из некоторого подмножества, задаваемого системой дополнительных уравнений и неравенств.

Методы оптимизации приводят к детерминированным рекуррентным алгоритмам оптимизации, т. е. решение задачи отыскивается в результате конечного числа итераций путем последовательного приближения к оптимальному, причем на каждом шаге используются новые выборочные значения обрабатываемых сигналов и параметров. Алгоритмы данного класса должны сходиться за конечное время к точке оптимума либо попадать в ее окрестность.

Следовательно, в результате использования итерационных алгоритмов получают последовательность значений искомого вектора (например, вектора коэффициентов цифрового фильтра), для которой значения функционала качества F отвечают соотношениям в случае минимизации (спуска) и в случае минимизации (спуска)

$$F\left[H(0)\right] > F\left[H(1)\right] > \ldots > F\left[H(n)\right] > \ldots$$

и в случае максимизации (подъема).

$$F\left[H(0)\right] < F\left[H(1)\right] < \ldots < F\left[H(n)\right] < \ldots$$

Точка H(0) определяет начальные условия процесса оптимизации.

Основными характеристиками итерационных алгоритмов являются сходимость (в математическом смысле) и скорость сходимости к оптимальному решению. Желательно, чтобы за меньшее время (количество итераций) выбранный алгоритм сходился к точке экстремума.

Процесс адаптации весового вектора H при использовании любого итерационного алгоритма может быть приведен к виду:

$$H(n+1) = H(n) + \mu(n)\bar{h}(n) \qquad (36)$$

где: μ(n) - величина шага итерации; h (n) - вектор, определяющий направление этого шага.

Варьируя процедуры выбора μ(n) и h (n), можно изменять методы поиска экстремума. Направление последующего шага логично выбирать так, чтобы осуществлялось приближение к оптимальному решению.

В зависимости от способа определения μ(n) и h(n) методы адаптации систем обработки сигналов можно разделить на три категории:

1. Методы прямого поиска, использующие только чистые значения функционала качества.
2. Методы, использующие, кроме того, первые производные функционала качества.
3. Методы, дополнительно использующие вторые производные функционала качества.

Методы прямого поиска применимы в случаях, когда функционал качества не задан в явном виде и определение производных затруднено, имеются точки разрыва рабочей функции, наблюдается несколько локальных экстремумов. Эти методы достаточно просты, но не всегда обеспечивают сходимость за конечное число шагов. Методы третьей категории обычно приводят к оптимальным решениям за более короткое время по сравнению с другими, однако они сложнее в алгоритмической реализации и не всегда обеспечивают ее устойчивость при неточном определении оцениваемых параметров. Наибольшее применение в адаптивной обработке сигналов нашли методы второй категории, основу которой в силу относительной простоты реализации составляют градиентные методы поиска экстремума.

Градиентные методы адаптации.

Градиентным называется метод, при котором каждый последующий вектор H(n + 1) выбирается в направлении -VF[H(n)], где VF[H(n)] - вектор-столбец частных производных функционала качества, называемый градиентом:

$$\nabla F\left[H(n)\right] = \begin{bmatrix} \dfrac{dF}{dh_0} \\ \dfrac{dF}{dh_1} \\ \vdots \\ \dfrac{dF}{dh_{N-1}} \end{bmatrix}. \qquad (37)$$

Таким образом, математическое выражение алгоритма адаптации принимает вид:

$$H(n+1) = H(n) - \mu\nabla F\left[H(n)\right]. \qquad (38)$$

Положительная константа μ определяется из условий устойчивости и требуемого времени сходимости алгоритма. Ее значение может быть переменным в зависимости от степени приближения текущего решения к оптимальному.

Рассмотрим свойства градиентного метода, определив квадратичный функционал качества в виде:

$$F(H) = H^T RH \,, \tag{39}$$

где R - положительно определенная симметричная матрица. В некоторых алгоритмах она может иметь смысл автокорреляционной матрицы конечной выборки отсчетов входного сигнала. Следовательно, F(H)>0 при любых H и достигает минимума при H = 0. Подобное тривиальное решение неприемлемо в задачах обработки сигналов, т. к. вырождает любой адаптивный цифровой фильтр [7].

Градиент функционала качества (39) равен:

$$\frac{dF(H)}{dH^T} = RH \,,$$

тогда в соответствии с (38)

$$H(n+1) = H(n) - \mu RH(n) \,. \tag{40}$$

Расписывая равенство (40) по итерационным шагам при любом начальном векторе H(0), получаем:

$$H(1) = H(0) - \mu RH(0) = (I - \mu R)H(0) \,,$$

$$H(2) = H(1) - \mu RH(1) = (I - \mu R)H(1) = (I - \mu R)^2 H(0)$$

$$\ldots$$

$$H(n) = (I - \mu R)^n H(0) \,. \tag{41}$$

Из формулы (41) следует, что при условии

$$I - \mu R = 0$$

оптимальное значение $H_o = 0$ может быть достигнуто за один шаг в направлении -VF[H(0)]. Для выполнения этого условия необходимо, чтобы

$$I = \mu R \quad \text{или} \quad R = \mu^{-1} I \,,$$

т. е. матрица R должна быть диагональной с элементами на главной диагонали. В противном случае для достижения оптимального решения потребуется большее число итераций.

Если начальную точку H(0) выбирать не произвольно, а определенным образом, то возможен другой способ достижения оптимального решения за один шаг. Для этого первое равенство в (41) перепишем в виде:

$$H(1) = H(0) - \mu RH(0) \,.$$

Если оптимальное значение $H_o = 0$ достигается за один шаг, т.е. H(1) = H0 = 0, то H(0) = μRH (0). Это возможно, если вектор H(0) является собственным вектором матрицы R, соответствующим собственному числу λ = μ. В этом случае соблюдается равенство:

$$H(1) = H(0) - \mu\mu^{-1} H(0) = 0 \,,$$

доказывающее одношаговую сходимость:

Таким образом, из условия

$$RH(0) = \mu^{-1} H(0) \tag{42}$$

следует, что оптимальное решение возможно найти за один шаг адаптации Для этого необходимо значение шага μ выбрать обратным одному из собственных чисел матрицы R, а H(0) - равным соответствующему собственному вектору. При этом матрица R может не быть диагональной.

Сходимость градиентных методов адаптации обеспечивается, если линии уровня функционала качества, соответствующие условию F = const, замкнуты вокруг точки экстремума при положительной определенности матрицы R. В противном случае последовательность точек итеративных решений не всегда сходится.

Выводы. Рекурсивность придает новые свойства алгоритму фильтрации, однако также накладывает некоторые ограничения на параметры настройки, которые не встречались у обычных фильтров. Возможна ситуация, когда входной сигнал фильтра становится корневым, то есть не изменяется при прохождении через фильтр. Очевидно, что такой эффект может привести к некоторой потере информации. Было обнаружено, что с помощью соответствующего подбора глубины рекурсии и параметров настройки взвешенной нормы данный эффект можно подавить. Так значения параметра $r > 0.5$ значительно ухудшают результат фильтрации. Рекомендуется выбирать глубину рекурсии в диапазоне $r \in [0.3, 0.5]$. Например, $r = 0.5$

Все это позволяет сделать вывод, что предлагаемая модификация медианного фильтра позволяет найти разумный компромисс в известном противоречии между степенью фильтрации шума и сохранением контрастных деталей обрабатываемого изображения.

1. Писаревский А.Н. Системы технического зрения. – Л.: Машиностроение, 1998. – 268 с.
2. http://www.aldis.ru/products
3. http://www.aldis.ru/omron/products
4. http://www.fastvideo.ru
5. Солонина А.И. Основы цифровой обработки сигналов / А.И. Солонина, С.М. Арбузов, Е.Б. Соловьева - СПб.: БХВ-Петербург, 2005. – 768 с.
6. Сергиенко А.Б. Цифровая обработка сигналов / А.Б. Сергиенко.- К.: ЦУЛ, 2004. – 300с.
7. Оппенгейм Э. Применение цифровой обработки сигналов /Є. Оппенгейм. – М.: Мир, 1980. – 312 с.